An overview of EMMA—Extensible MultiModal Annotation

Michael Johnston
AT&T Labs Research
8/9/2006
Overview

• Introduction to EMMA
  – EMMA Structural elements and Attributes
  – Examples of EMMA annotation

• Support in EMMA for integrating inputs from multiple modes
  – emma:hook
Introduction to EMMA

- Spoken and multimodal interactive systems involve communication among different components
  - Recognition (Speech, Handwriting, Gestures)
  - Understanding
  - Multimodal Integration
  - Dialog/Interaction management
Spoken Input Processing

SPEECH RECOGNITION

“flight from denver to boston tomorrow”

UNDERSTANDING

<flight>
<source>denver</source>
<dest>boston</dest>
<time>tomorrow</time></flight>

DIALOG MANAGER
Multimodal Input Processing

SPEECH RECOGNITION

“zoom in here”

UNDERSTANDING

<command><action>zoom</action>
<location><type>area</type><points>42.1345 -37.128 ... </points></location><time>tomorrow</time></command>

GESTURE RECOGNITION

area ...

UNDERSTANDING

<location><type>area</type><points>42.1345 -37.128 ... </points></location>

MULTIMODAL INTEGRATION

<command><action>zoom</action>
<location><type>area</type><points>42.1345 -37.128 ... </points></location><time>tomorrow</time></command>

DIALOG/INTERACTION MANAGER
Introduction to EMMA (cont.)

• EMMA (Extensible MultiModal Annotation)
  – Standardized XML markup language for communication among components of spoken and multimodal interactive systems
  – Provides structural containers for possible interpretations of user inputs and elements and attributes for annotating properties of user inputs (time, confidence, input source etc.)
  – Critically, does not standardize the representation of interpretations of user input
Spoken Input Processing

SPEECH RECOGNITION

EMMA

UNDERSTANDING

EMMA

DIALOG MANAGER
Multimodal Input Processing

SPEECH RECOGNITION

GESTURE RECOGNITION

UNDERSTANDING

MULTIMODAL INTEGRATION

DIALOG/INTERACTION MANAGER
Introduction to EMMA (cont.)

- EMMA (Extensible MultiModal Annotation)
  - Replacement for NLSML
  - Aims to facilitate interoperation among modality components from different vendors
  - Will also play a role in logging and corpus annotation
EMMA Standards Effort at W3C
- [http://www.w3.org/TR/emma](http://www.w3.org/TR/emma)

- EMMA is part of W3C standards effort in the Multimodal Interaction (MMI) Working Group
  ([http://www.w3.org/2002/mmi](http://www.w3.org/2002/mmi))
- EMMA requirements published on Jan. 11, 2003
- First EMMA working draft published on August 11, 2003
- Last call working draft September 2005
  - [http://www.w3.org/TR/emma/](http://www.w3.org/TR/emma/)

EMMA Structural Elements

• Provide containers for application semantics and for multimodal annotation

- E.g <emma:emma ...>
  <emma:one-of>
    <emma:interpretation>
      ....
    </emma:interpretation>
    <emma:interpretation>
      ....
    </emma:interpretation>
  </emma:one-of>
</emma:emma>

EMMA Elements

<table>
<thead>
<tr>
<th>emma:emma</th>
</tr>
</thead>
<tbody>
<tr>
<td>emma:interpretation</td>
</tr>
<tr>
<td>emma:one-of</td>
</tr>
<tr>
<td>emma:group</td>
</tr>
<tr>
<td>emma:sequence</td>
</tr>
<tr>
<td>emma:lattice</td>
</tr>
</tbody>
</table>
EMMA Annotations

- Characteristics and processing of input, e.g.:

<table>
<thead>
<tr>
<th>emma:tokens</th>
<th>token of input</th>
</tr>
</thead>
<tbody>
<tr>
<td>emma:process</td>
<td>reference to processing</td>
</tr>
<tr>
<td>emma:no-input</td>
<td>lack of input</td>
</tr>
<tr>
<td>emma:uninterpreted</td>
<td>uninterpretable input</td>
</tr>
<tr>
<td>emma:lang</td>
<td>human language of input</td>
</tr>
<tr>
<td>emma:signal</td>
<td>reference to signal</td>
</tr>
<tr>
<td>emma:media-type</td>
<td>media type</td>
</tr>
<tr>
<td>emma:confidence</td>
<td>confidence scores</td>
</tr>
<tr>
<td>emma:source</td>
<td>annotation of input source</td>
</tr>
<tr>
<td>emma:start emma:end</td>
<td>Timestamps (absolute/relative)</td>
</tr>
<tr>
<td>emma:medium emma:mode emma:function</td>
<td>medium, mode, and function of input</td>
</tr>
<tr>
<td>emma:hook</td>
<td>hook</td>
</tr>
</tbody>
</table>
Examples of EMMA Annotation

<emma:emma version="1.0" xmlns="http://www.w3.org/2003/04/emma">
  <emma:interpretation id="interp1"
    emma:start="1087995961542"
    emma:end="1087995963542"
    emma:tokens="flights from boston to denver"
    emma:confidence="0.6"
    emma:medium="acoustic"
    emma:mode="speech"
    emma:function="dialog"
    emma:source="http://example.com/microphone/NC-61"
    emma:signal="http://example.com/signals/sg23.wav"
    emma:media-type="audio/dsr-202212; rate:8000; maxptime:40">
    <origin>Boston</origin>
    <destination>Denver</destination>
  </emma:interpretation>
</emma:emma>
Examples of EMMA Annotation

- **Use of EMMA elements**: `<emma:one-of>`, `<emma:interpretation>`

```
<emma:emma version="1.0"
   xmlns:emma="http://www.w3.org/2003/04/emma">
  <emma:one-of id="r1">
    <emma:interpretation id="int1" emma:confidence="0.75"
                        emma:tokens="flights from boston to denver">
      <origin>Boston</origin>
      <destination>Denver</destination>
    </emma:interpretation>
    <emma:interpretation id="int2" emma:confidence="0.68"
                        emma:tokens="flights from austin to denver">
      <origin>Austin</origin>
      <destination>Denver</destination>
    </emma:interpretation>
  </emma:one-of>
</emma:emma>
```
Multimodal Integration: emma:hook

• One of the most powerful aspects of multimodal interfaces is the possibility of providing support for user inputs optimally distributed over the available input modes e.g. **SPEECH**: “zoom in here” **INK:**
Multimodal Integration: emma:hook

- In order to support composite multimodality some kind of multimodal integration mechanism is needed
  - Checks compatibility of content from different modes and combines them
- The integration mechanism is not standardized
- EMMA provides an annotation **emma:hook** which can be used to mark locations in the semantic representation where content from another mode is needed
- **emma:hook** can be used to drive different kinds multimodal integration mechanisms
Multimodal integration: emma:hook

• The value of emma:hook indicates the mode of the content to be integrated
  – emma:hook="ink"
  – <location emma:hook="ink"/>
    indicates that this element needs to be combined with a
    <location> element from the ink mode
      • <emma:interpretation emma:mode='ink'>
        – <location>
          ....
        </location>
Multimodal integration: emma:hook

SPEECH:

“ZOOM IN HERE”

<emma:interpretation emma:mode="speech">
<command>
  <action>zoom</action>
  <location emma:hook="ink" >
    <type>area</type>
  </location>
</command>
</emma:interpretation>
Multimodal integration: emma:hook

Speech interpretation can be generated using SRGS rules from the W3C SI specification.

```
<rule id="zoom"> zoom in here
<tag> $.command = new Object();
$.command.action = "zoom";
$.command.location = new Object();
$.command.location._attributes = new Object();
$.command.location._attributes.hook = new Object();
$.command.location._attributes.hook._nsprefix = "emma";
$.command.location._attributes.hook._value = "ink";
$.command.location.type = "area";
</tag>
</rule>
```
Multimodal integration: emma:hook

**SPEECH:**

“ZOOM IN HERE”

**INK:**

![Map with a circle drawn around a location]

```xml
<emma:interpretation emma:mode="speech">
  <command>
    <action>zoom</action>
    <location emma:hook="ink">
      <type>area</type>
    </location>
  </command>
</emma:interpretation>

<emma:interpretation emma:mode="ink">
  <location>
    <type>area</type>
    <points>42.1345 -37.128 42.1346 37.120 ...</points>
  </location>
</emma:interpretation>
```
Multimodal integration: emma:hook

SPEECH: “ZOOM IN HERE”
<emma:interpretation emma:mode="speech">
  <command>
    <action>zoom</action>
    <location emma:hook="ink">
      <type>area</type>
    </location>
  </command>
</emma:interpretation>

INK: (CIRCLE ON MAP)
<emma:interpretation emma:mode="ink">
  <location>
    <type>area</type>
    <points>42.1345 -37.128 42.1346 37.120 ...</points>
  </location>
</emma:interpretation>

MULTIMODAL: “ZOOM IN HERE” + CIRCLE
<emma:interpretation emma:mode="multiple">
  <command>
    <action>zoom</action>
    <location>
      <type>area</type>
      <points>42.1345 -37.128 42.1346 37.120 ...</points>
    </location>
  </command>
</emma:interpretation>
Conclusion

• EMMA: Extensible MultiModal Annotation markup language

• Provides standard representation for communication among the components that make up spoken and multimodal systems
  – Does not standardize application semantic representation
  – Provides
    • Containers for application specific semantics
    • Standard set attributes and elements for metadata
  – Enables interoperability among components from different vendors
  – Built in ‘hooks’ for multimodal integration (emma:hook)
Empowering Your customers and employees with speech technologies

SpeechTEKI 2006
Thank You